

# IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data

Joanna L. Sharman<sup>1</sup>, Chidochangu P. Mpamhanga<sup>1</sup>, Michael Spedding<sup>2</sup>,  
Pierre Germain<sup>3</sup>, Bart Staels<sup>4</sup>, Catherine Dacquet<sup>2</sup>, Vincent Laudet<sup>5</sup>,  
Anthony J. Harmar<sup>1,\*</sup> and NC-IUPHAR<sup>†</sup>

<sup>1</sup>University/BHF Centre for Cardiovascular Science, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, UK, <sup>2</sup>Institut de Recherches Internationales Servier, 92150 Suresnes, <sup>3</sup>Centre de Biochimie Structurale, Centre National de la Recherche Scientifique, 34090 Montpellier, <sup>4</sup>Université Lille Nord de France, Inserm U1011, Institut Pasteur de Lille, 59000 Lille and <sup>5</sup>Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, 69364 Lyon, France

Received September 10, 2010; Revised October 12, 2010; Accepted October 13, 2010

## ABSTRACT

The IUPHAR database is an established online reference resource for several important classes of human drug targets and related proteins. As well as providing recommended nomenclature, the database integrates information on the chemical, genetic, functional and pathophysiological properties of receptors and ion channels, curated and peer-reviewed from the biomedical literature by a network of experts. The database now includes information on 616 gene products from four superfamilies in human and rodent model organisms: G protein-coupled receptors, voltage- and ligand-gated ion channels and, in a recent update, 49 nuclear hormone receptors (NHRs). New data types for NHRs include details on co-regulators, DNA binding motifs, target genes and 3D structures. Other recent developments include curation of the chemical structures of approximately 2000 ligand molecules, providing electronic descriptors, identifiers, link-outs and calculated molecular properties, all available via enhanced ligand pages. The interface now provides intelligent tools for the visualization and exploration of ligand structure-activity relationships and the structural diversity of compounds active at each target. The database is freely available at <http://www.iuphar-db.org>.

## INTRODUCTION

Complementary to the familiar large-scale data banks holding the outputs of genomics research are many

smaller, highly focused databases, whose remit is the comprehensive curation of selected subsets of genes through annotation of historic and emerging biological literature. Such resources are often created to address the specific needs of particular research fields and normally rely on the voluntary efforts of interested experts and a community of scientists. These resources can provide distilled data tailored to the requirements of the end user. The need for such facilities will only intensify as the amount and complexity of scientific data increases.

In 2009 we reported the development of an expert-curated database (1) from the International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR, <http://www.iuphar.org/nciuphar.html>). The IUPHAR database (IUPHAR-DB, <http://www.iuphar-db.org>) displays detailed information on the structure, function, anatomy, pathophysiology and pharmacology of G protein-coupled receptors (GPCRs), voltage-gated ion channels (VGICs) and ligand-gated ion channels (LGICs). Members of these protein superfamilies correspond to the biological targets of at least one-third of licensed medicinal drugs (2).

The database is a result of the voluntary efforts of over 700 international expert scientists coordinated by NC-IUPHAR through its network of subcommittees. The content and curation model of IUPHAR-DB have already been described (1); here we will focus on advances in the construction, content and user interface. We announce the addition of a fourth class of important drug targets and transcription regulators, the nuclear hormone receptors (NHRs). Furthermore, we provide details on improvement to the curation process for chemical compounds, which has resulted in the first

\*To whom correspondence should be addressed. Tel: +44 131 242 6693; Fax: +44 131 242 6782; Email: [tony.harmar@ed.ac.uk](mailto:tony.harmar@ed.ac.uk); [tharmar@mac.com](mailto:tharmar@mac.com)

<sup>†</sup>The members of NC-IUPHAR group are provided in Acknowledgements.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

rigorous definition of chemical entities in IUPHAR-DB. We also introduce a range of new features that enhance usability of the database, including redesigned ligand pages and intelligent tools for structure-based exploration of ligand–target interactions. We hope these will allow interested chemists and pharmacologists to rapidly visualize and appreciate the diversity of ligands, their chemical profiles and the structure activity relationships represented in the data.

## CONTENT

In its current release the database contains information on proteins encoded by 356 GPCR genes, 141 VGIC subunit genes, 52 LGIC subunit genes and 49 NHR genes in human, rat and mouse. The database documents 6533 interactions between these proteins and 2715 distinct ligand molecules. All data included in IUPHAR-DB are curated from the primary literature by expert subcommittees or individual experts appointed by NC-IUPHAR or by in-house curators and then reviewed by external experts. The database currently references 11 332 unique publications, 99% of which are linked to PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). Supplementary Figure S1 provides charts describing the content of gene families in IUPHAR-DB.

## NEW FEATURES

### Nuclear hormone receptors

Recently, the database has expanded to provide the first comprehensive annotation of the literature on the pharmacology of 49 human NHR gene products and their rodent orthologues, together with relevant information on their structure and function. The information in IUPHAR-DB complements that available in other online resources such as the Nuclear Receptor Signaling Atlas (NURSA) (3), an online resource with curated information and experimental data for receptors, ligands and co-regulators and NureXbase (<http://nurexbase.prabi.fr>), a multi-species genomic database with structural information for NHRs, complexes and endocrine disruptors.

Unlike the other receptors in IUPHAR-DB, which are cell surface proteins that transduce extracellular signals by modulating intracellular concentrations of second messengers or ions, NHRs are intracellular, non-transmembrane proteins with distinct structural and functional properties (4). NHRs are transcriptional regulators that exert their actions by binding to sequence-specific promoter elements on target genes. NHR ligands include steroid and thyroid hormones, metabolites and xenobiotics; many NHRs are ‘orphan’ receptors, for which the endogenous ligands are still unknown or may not exist.

IUPHAR-DB provides details of recommended nomenclature, synonyms used in the literature, structure, genomic location, physiological function, tissue distribution, functional assays, biologically significant alternatively spliced and RNA-edited variants, the clinical-relevance of polymorphisms and phenotypes resulting from alterations in gene-expression. Pharmacological information

includes the affinities of selected agonists and antagonists, such as approved drugs, commonly used experimental compounds and endogenous ligands. For NHRs, additional ‘custom’ information is provided, including details of consensus DNA response elements, co-binding partners required for DNA binding, co-repressor and co-activator proteins and the main target genes regulated by NHRs. Each NHR in IUPHAR-DB is hyperlinked to the corresponding entries in other relevant databases, including several of the NCBI Entrez databases (5), the human, rat and mouse genome databases [HGNC (6), MGI (7) and RGD (8)] and NURSA. Links to other databases containing information on drug targets and their ligands are provided on a ‘Useful Links’ page, which is regularly updated.

### Three-dimensional structures

Structural information about receptors and channels has been enhanced by the addition of links to experimentally solved protein structures in the RCSB Protein Data Bank (RCSB PDB) (9). There is extensive three dimensional (3D) structural information available on NHRs, so for each receptor we have selected at least three representative structures, including, where available, an ‘apo’ structure (protein with no ligand bound), a protein bound to an agonist and a protein bound to an antagonist. Details have been added to the database for the few GPCR, VGIC and LGIC structures that are known. Database pages also include dynamic links to complete lists of structures available in the RCSB PDB for each protein and from IUPHAR-DB ligands to PDB ligand pages.

### Ligand classification and structural information

Previously, information on ligand molecules in IUPHAR-DB was limited to common names and synonyms and PubChem compound identifiers (CIDs) (10). In 2009 we began an initiative to curate and classify all ligand molecules in IUPHAR-DB, incorporating structural specification, computational molecular descriptors and other information. Rigorous definition of chemical compounds brings many advantages to both database developers and users; for instance knowledge of chemical structure reduces unintended duplication where multiple publications refer to the same compound under different synonyms or fail to provide the chemical structure. It is now possible to unambiguously identify compounds in IUPHAR-DB and use molecular descriptors to search for them in other databases. We have extended the functionality of our in-house curation tools to enable straightforward curation of chemical structures.

During the chemical curation process we manually checked each compound against the primary literature to ensure that chemical structures were attributed to the correct compound names, removed over 200 duplicates and associated each compound with PubChem CIDs where possible. In all cases we endeavoured to provide correct isomeric forms; however, we found that stereomeric information is often omitted in the

pharmacological literature. Consequently, we were obliged to use the non-isomeric representations for all cases where enantiomers or diastereomers are not specified in the primary assay used; this reflects some of the unavoidable historical ambiguity seen in pharmacological data.

About two-thirds of the IUPHAR-DB compounds were unambiguously identified in the PubChem compounds database. Utilizing the PubChem download service we were able to retrieve 1944 chemical structures in simplified molecular input line entry specification (SMILES) format (<http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>). To enable the full and unambiguous identification of IUPHAR compounds, each molecular entity is represented by its common name, a list of synonyms, a systematic name, unique chemical identifiers and several structural descriptors and formats. The Open Babel software package ([http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)) was used to unify the format of the retrieved SMILES strings and create various structural descriptors. Generated descriptors include two types of SMILES descriptor, one containing isomeric specification of chiral compounds and a second without chiral or isotopic specification, IUPAC International Chemical Identifiers (InChI) and InChI Keys (<http://www.iupac.org/inchi>). Systematic IUPAC names and two-dimensional (2D) images for use on the IUPHAR-DB website were generated using the NCI/CADD Chemical Identifier Resolver from the National Cancer Institute (<http://cactus.nci.nih.gov/chemical/structure>). Various physico-chemical properties were calculated using the Chemistry Development Kit (CDK) (11), including the five Lipinski 'drug-likeness' measures (12): polar surface area (TPSA) (13), predicted LogP (XLogP) (14), molecular weight, number of hydrogen bond donors and acceptors. The number of rotatable bonds is also included because this can be a useful indication of the compound flexibility and complexity.

The final set of 2715 distinct compounds were then classified into the following classes: synthetic organic-based compounds, inorganic molecules (e.g. ions), simple naturally occurring bioactive compounds, natural products, peptides (both synthetic and naturally occurring) and an 'other' class for the few compounds that do not fall into any of the former categories. Supplementary Figure S1 describes the distribution of IUPHAR-DB ligands into their classes.

### Enhanced ligand pages

Each ligand entry in IUPHAR-DB has a dedicated data page displaying information about its structural properties and biological activity (Figure 1). Where available, an image of the ligand's 2D structure and physicochemical properties are displayed. A series of tabs provide access to further information. A summary tab includes the ligand's classification, whether it is known to be an approved drug, the systematic name and other synonyms. To help users find further information on licensed drugs and experimentally used compounds we provide links to databases such as DrugBank (15), the

Pharmacogenomics Knowledge Base (PharmGKB) (16) and Chemical Entities of Biological Interest (ChEBI) (17).

A second tab contains information on the ligand's biological activity with tables listing its selectivity at receptors in IUPHAR-DB and links to the relevant receptor pages. A third tab displays a reference list and a fourth tab provides descriptors and download options for the chemical structure. Another tab provides a display of any structurally similar compounds that exist in the database, with lists of their receptor targets and links to the relevant data pages. A useful feature is the ability to quickly launch a chemical editor (by clicking on the ligand's image or following a link under the 'Similar compounds' tab), allowing the structure to be modified and used for structure-based searching of the database (see below for details). Pharmacological data tables on receptor pages have also been enhanced with structural information about ligands, which helps users to visualize the types of compounds that interact with specific receptors.

### New search tools

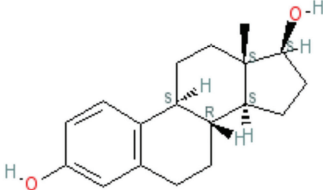
In addition to text-based searches of the database content the search tools now allow receptor and ligand retrieval by external identifier, such as by protein accession or DrugBank identifier. Moreover, users are now able to navigate ligand structure space by drawing a chemical structure or pasting a SMILES string into a chemical editor [ChemAxon's MarvinSketch (<https://www.chemaxon.com/marvin/sketch/index.php>) Java applet] and performing a structure-based query of the database. The available search methods are substructure, exact, similarity and SMARTS-pattern (<http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>) matching. Structure-based searches are powered by Dotmatics' Pinpoint software ([http://www.dotmatics.com/products\\_pinpoint.jsp](http://www.dotmatics.com/products_pinpoint.jsp)). On the results page users see the structures of matched compounds and a quick overview of their receptor targets, with links to ligand and receptor pages and options to refine the search such as limiting by molecular weight or polar surface area (Figure 2).

### FUTURE PERSPECTIVES

While we have obtained chemical structures for the majority of small molecule ligands in IUPHAR-DB, structural information for a majority of the approximately 900 endogenous and synthetically derived peptide ligands is lacking. Furthermore, the implemented structure-based search methods are not appropriate for oligopeptides, for which sequence-based methods may be preferable. In a future database release we aim to provide primary sequences and information on the post-translational modification of endogenous peptides and also the amino acid sequences and chemical modifications of synthetic peptides. We also plan to expand the coverage of IUPHAR-DB to include further protein families and new data types.

<b>IUPHAR-DB Ligand: 1013</b>	
Ligand name	<b>17<math>\beta</math>-estradiol</b>

<b>2D Structure</b> <span>?</span> 	<b>Calculated Physical-Chemical Properties</b> <span>?</span> <table> <tr> <td>Hydrogen bond acceptors</td><td>1</td></tr> <tr> <td>Hydrogen bond donors</td><td>2</td></tr> <tr> <td>Rotatable bonds</td><td>0</td></tr> <tr> <td>Topological polar surface area</td><td>40.46</td></tr> <tr> <td>Molecular weight</td><td>272.18</td></tr> <tr> <td>XLogP</td><td>4.65</td></tr> <tr> <td>No. Lipinski's rules broken</td><td>0</td></tr> </table> <p>Molecular properties generated using the <a href="#">CDK</a></p>	Hydrogen bond acceptors	1	Hydrogen bond donors	2	Rotatable bonds	0	Topological polar surface area	40.46	Molecular weight	272.18	XLogP	4.65	No. Lipinski's rules broken	0
Hydrogen bond acceptors	1														
Hydrogen bond donors	2														
Rotatable bonds	0														
Topological polar surface area	40.46														
Molecular weight	272.18														
XLogP	4.65														
No. Lipinski's rules broken	0														

<b>Summary</b>	<b>Biological activity</b>	<b>References</b>	<b>Structure</b>	<b>Similar compounds</b>
----------------	----------------------------	-------------------	------------------	--------------------------

<b>Selectivity at human GPCRs</b>						
<b>Key to terms and symbols</b>				Click column headers to sort		
Receptor	Type	Action	Affinity	Units	Concentration range (M)	Reference
<a href="#">GPER</a>	Agonist	Full agonist	8.2 – 8.5	pK <sub>i</sub>	-	<a href="#">1-2</a>

<b>Selectivity at human ion channels</b>						
<b>Key to terms and symbols</b>				Click column headers to sort		
Receptor	Type	Action	Affinity	Units	Concentration range (M)	Reference
<a href="#">K<sub>Ca</sub>1.1</a>	Activator	None	5.6	pEC <sub>50</sub>	-	<a href="#">3</a>

<b>Selectivity at human nuclear hormone receptors</b>						
<b>Key to terms and symbols</b>				Click column headers to sort		
Receptor	Type	Action	Affinity	Units	Concentration range (M)	Reference
<a href="#">Estrogen receptor-<math>\alpha</math></a>	Agonist	Agonist	9.82	pK <sub>i</sub>	-	<a href="#">4</a>
<a href="#">Estrogen receptor-<math>\beta</math></a>	Agonist	Agonist	9.34	pK <sub>i</sub>	-	<a href="#">4</a>

**Figure 1.** Part of a ligand page showing an image of the 2D structure, calculated physicochemical properties and tables of biological activity at GPCRs, ion channels and NHRs.



Your query returned 4 matches:

Order results by: Topological polar surface area Limit results by: None 0.0 Update

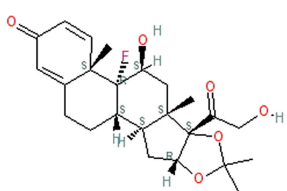
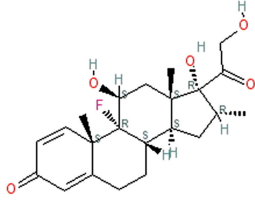
 <p>Name: <a href="#">triamcinolone acetonide</a> Targets: <a href="#">Glucocorticoid receptor</a> <a href="#">Use in search</a></p>	 <p>Name: <a href="#">dexamethasone</a> Targets: <a href="#">Glucocorticoid receptor</a>, <a href="#">Mineralocorticoid receptor</a>, <a href="#">Pregnane X receptor</a> <a href="#">Use in search</a></p>
---	--

Figure 2. Part of the chemical structure search results page.

## CITING IUPHAR-DB

For a general citation of the resource we recommend citing this article. For citing specific receptor pages we recommend a format similar to the following example: A. P. Davenport, E. J. Mead. Kisspeptin receptor. Last modified on <date>. Accessed on <date>. IUPHAR database (IUPHAR-DB), <http://www.iuphar-db.org/DATABASE/FamilyMenuForward?familyId=34>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Daniel Ormsby at Dotmatics Limited for the Pinpoint license. The authors are grateful to Peter Buneman and Heiko Mueller at the University of Edinburgh for database expertise. The authors especially thank all contributors and members of NC-IUPHAR and its subcommittees for their ongoing support. The authors also acknowledge the support of the British Heart Foundation Centre of Research Excellence Award (RE/08/001). International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR) members: T.I. Bonner, W.A. Catterall, A.P. Davenport, P. Delagrè, C.T. Dollery, S. Duckles, S. Enna, S.M. Foord, P. Germain, A.J. Harmar, V. Laudet, G. Milligan, R.R. Neubig, E.H. Ohlstein, J. Peters, J.P. Pin, U. Ruegg, D.B. Searls, M. Spedding and M.W. Wright.

## FUNDING

British Pharmacological Society, Abbott, GlaxoSmithKline, Incyte, Millipore, Novartis, Servier,

UNESCO and Wyeth. Funding for open access charge: Waived by Oxford University Press.

*Conflict of interest statement.* None declared.

## REFERENCES

- Harmar, A.J., Hills, R.A., Rosser, E.M., Jones, M., Buneman, O.P., Dunbar, D.R., Greenhill, S.D., Hale, V.A., Sharman, J.L., Bonner, T.I. *et al.* (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res.*, **37**, D680–D685.
- Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1**, 727–730.
- McKenna, N.J., Cooney, A.J., DeMayo, F.J., Downes, M., Glass, C.K., Lanz, R.B., Lazar, M.A., Mangelsdorf, D.J., Moore, D.D., Qin, J. *et al.* (2009) Minireview: Evolution of NURSA, the Nuclear Receptor Signaling Atlas. *Mol. Endocrinol.*, **23**, 740–746.
- Germain, P., Staels, B., Dacquet, C., Spedding, M. and Laudet, V. (2006) Overview of nomenclature of nuclear receptors. *Pharmacol. Rev.*, **58**, 685–704.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–D16.
- Bruford, E.A., Lush, M.J., Wright, M.W., Sneddon, T.P., Povey, S. and Birney, E. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.*, **36**, D445–D448.
- Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E. and Blake, J.A. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
- Twigger, S.N., Shimoyama, M., Bromberg, S., Kwitek, A.E. and Jacob, H.J. (2007) The Rat Genome Database, update 2007—easing the path from disease to data and back again. *Nucleic Acids Res.*, **35**, D658–D662.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J. and Bryant, S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–W633.
- Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R. and Willighagen, E.L. (2006) Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.*, **12**, 2111–2120.
- Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.
- Ertl, P., Rohde, B. and Selzer, P. (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.*, **43**, 3714–3717.
- Wang, R., Fu, Y. and Lai, L. (1997) A new atom-additive method for calculating partition coefficients. *J. Chem. Inform. Comput. Sci.*, **37**, 615–621.
- Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B. and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
- Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E. *et al.* (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J.*, **1**, 167–170.
- de Matos, P., Alcantara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S. and Steinbeck, C. Chemical Entities of Biological Interest: an update. *Nucleic Acids Res.*, **38**, D249–D254.